THE SPATIAL DEMAND FOR OFFICE SPACE IN BOSTON

By

Michael Z. Allen

B. S., Economics
B. S., Geography
The Ohio State University, 1996

Submitted to the Department of Urban Studies and Planning
in Partial Fulfillment of the Requirements for the Degree of

MASTER IN CITY PLANNING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1998

Signature of Author..
Michael Z. Allen
May 20, 1998

Certified by ....
William C. Wheaton
Professor of Economics

Accepted by ....
Lawrence S. Bacow
Professor of Law and Environmental Policy
Chair, Master in City Planning Committee

The Spatial Demand for Office Space in Boston

by

Michael Z. Allen

Submitted to the Department of Urban Studies and Planning on May 15, 1998 in partial fulfillment of the requirements for the degree of Master of City Planning

ABSTRACT

Ordinary Least Square regression combined with GIS data analysis was used to test hypotheses seeking to explain the spatial variation in the average asking rent for office space in metropolitan Boston. The model priced the constituent elements of the locational aspect of office rents such that a continuous surface of potential rent was derived. The model is further used to validate current theories of urban economics regarding local workers, commuting, and agglomeration economies. The model of spatial demand developed in the thesis is intended to provide both planners and developer/investors with a more accurate means with which to value land with respect to the office market.

The hedonic value analysis validated all of the traditional urban economic hypotheses put forth, namely that office firms value proximity to affluent neighborhoods, the labor force, other office firms, retail, and transportation infrastructure. The above variables were adapted to different spatial resolutions with the aid of a geographic information system (GIS) and tested for significance at each level. To best capture their intended effects, the locational variables were spatially reformulated to be context sensitive. This methodology is a departure from reliance on predefined geographies and is proposed as a way in which such studies can build a cumulative record comparable across studies.

This thesis presents a new application of GIS as an integrated analytical tool in econometric analysis in urban economics. However, several critical issues remain unresolved. Among them are the need for broader cross-section and longitudinal data, as well as more spatially sophisticated statistical techniques.

Thesis Supervisor: William C. Wheaton
Title: Professor of Economics

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

Problem Definition

In a free market economy, the primary signifier of value is price. Individual decisions simultaneously effect and are affected by changes in the price of a good. Accurate pricing information is thus crucial to efficient resource allocation decisions. The absence of efficient pricing leads not only to individual loss but economic (deadweight) loss and market inefficiencies. A main, if not the primary, component of the resource allocation decision in real estate is *where* to build or invest. In the case of commercial office space, the predicted rent of a building in a given location guides the decision to build or invest. The durable character of the built environment combined with the systematic lack of accurate property pricing produces a distorted urban form. For example, an optimal location for an office center might contain a regional mall in its place due to misinformation regarding what land use would yield the highest rent per unit area. While it is difficult to prove what this geographical distortion of space costs the economy, under the ideal conditions of perfect information and market pricing, the built environment would be better constructed to efficiently serve the economy. A better understanding of location and how it influences rents contributes to this end.

One way to better understand the spatial variation in value is to analyze statistically the rental price of office space for insight into its *locational* determinants. However, the research efforts in this field are difficult to compare for methodological reasons, preventing a *cumulative* understanding of the problem. The problem that this thesis addresses is thus both substantive and methodological.

Objectives and Research Questions

There are three main goals of this thesis. They consist of the development of a hedonic price model, the use of the model to evaluate current theories of urban economics, and lastly to produce, through that model, an analytically informed predicted surface of potential office rent for the entire study area. The first of the goals highlights a new methodology for a locationally sensitive pricing model of commercial office rent. The model is developed by breaking down location into elemental parts that can be explicitly identified and systematically quantified. Those elemental parts reflect several research questions in the field of urban economics as to the factors that comprise the elemental parts of location. These questions serve as the basis for an analytical approach to modeling office rents.

- Are office rents correlated with neighborhood income?
- Are office rents correlated with access to the labor market?
- Is there evidence of agglomeration/ urbanization economies in the office market?
- Are office rents correlated with access to retail opportunities?
- Are office rents correlated with access to transportation infrastructure?
- Are locational factors stable through time?

The model of office rents is used to calibrate a combined locational value with respect to a hypothetical office building. The final substantive objective is thus to derive a predicted value of office rents by location. In doing so, the surface demonstrates a rent potential for all locations, including and especially those currently devoid of office buildings. Furthermore, it is the object of the thesis to accomplish the above in via a *repeatable* methodology, such that future work can build upon a knowledge base of locational pricing.

Thesis Organization

The thesis is composed of 5 chapters, which are organized to explicate the underlying theory, methodology, and results of the research. Chapter Two describes the theoretical basis of the model as well as its fundamental assumptions. Within the discussion of theory incorporated into this thesis is a review of previous research. Reviewed are similar works that address the locational component of rent. However, these models fall short in their ability to incorporate location in fine detail. To improve upon previous research, a general form of a more spatially sensitive model is proposed. The chapter also includes a discussion of the sources of data used in both the exploratory and final models.

Chapter 3 explains in detail the methodology employed in the construction of the model. Two main techniques employed are econometric/hedonic modeling and the use of geographic information systems (GIS). Included in the section on GIS is a discussion of the process of creating and experimenting with spatial variables in the modeling process. The exploration of space facilitated by the GIS also led to the methodological decision to split the data into two regions. The chapter concludes with an explanation of how the dimension of time is managed and its implications regarding model estimation.

Chapter Four reports and interprets the results of the model, including the magnitude and meaning of each variable, its coefficient, as well as overall issues and problems. In summary, the results show that location has quantifiable value that can be modeled through estimation of its constituent parts. The hypotheses surrounding the major economic theories of firm location are answered in the affirmative and revealed significant, but the other access related results are mixed. The chapter also contains the predicted surface of potential rent along with a detailed explanation of the map.

The final chapter, Five, is the summary and thesis conclusion. It contains an elaboration of the model's shortcomings both theoretically and empirically. The research finds that a richer description of location is needed, which requires both enhanced data collection efforts as well as a more sophisticated theoretical framework.

CHAPTER 2
THEORY & LITERATURE REVIEW

Chapter Introduction

This chapter introduces as background the appropriate theories of urban economics on which this thesis is based. It begins with the traditional bid-rent theory of land value in a monocentric city, and continues with relevant aspects of the theories of agglomeration and decentralization. The discussion then moves into a review of previous research regarding empirical estimation of real estate property values. The discussion is specifically focused on the treatment of location in this literature and how this research presents a point of departure.

Theories of Urban Economics

Urban economics is the study of the spatial organization of labor and capital at particular points in space, specifically in cities [1, 19, 22]. As this thesis focuses on the locational value of office space, it is appropriate to begin with the traditional urban theory of land rent. The basic premise of Von Thunen's rent curve shown below is that a point , namely the center of the city, exhibits some kind of productive advantage [37].

BID-RENT CURVE

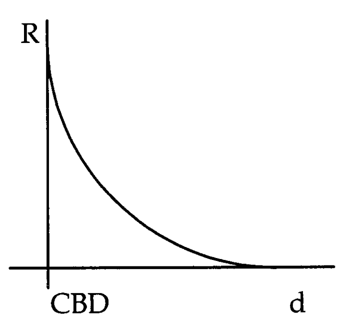


Figure 1.

All else equal, entities (people, firms, and governments) are willing to pay more to rent land (R) proximitous to the center (CBD). Under the competitive market assumptions on which the theory rests, the intrinsic productive value of a location is translated directly into rent. Thus, land rent is considered the residual of the productivity of a place [1, 11, 19, 25]. This thesis accepts this general relationship of the effects of a location based productive advantage, but asserts, in accordance with Sivitanidou [29], that the bid-rent model is not fully adequate to explain locational value due to the polycentric nature of contemporary cities. The analysis presented here assumes that the spatial variation in the rent of office space is evidence of the productive advantage of place (which need not necessarily be *the* center), but seeks to expand the economic literature on the determinants of productive advantage.

The theories of productive advantage that are appropriate to this research on office building rents are those which seek to explain the existence of commercial concentration within a general framework of urban decentralization. The specific theories relevant to this research are those of agglomeration economies and local workers, respectively. Included in the following review are contributions directly associated with office markets, as well as those from housing economics, in which there has been more such scholarship to date [36].

Within the framework of the market for office space, agglomeration economies are assumed to be the result of the premium placed on communication in the business world [21, 37]. Mun and Hutchinson show that the premium placed on face to face communication is proportional to distance [21]. This means that all else equal, firms are willing to pay more to be near other firms of like industries. Distance measures similar to those used in their paper are thus used in this research. This thesis, however, departs methodologically in its treatment of agglomeration by using disaggregate data, an approach which is fully discussed in Chapter 3.

Another theory of the productive advantage of location which stems directly from the traditional monocentric city of [1, 22] is that of the multi-centric city, or alternatively, that of decentralization. The aspect of decentralization used in this research is based on the theory of local workers [11, 19]. In the traditional urban model [1, 11, 19], employment is assumed to be wholly concentrated at the center of the city. The increase in workers' transportation costs with distance from the center is exactly offset by lower land rent costs. This relationship of constant combined costs makes the worker indifferent to location, ceterus paribus. Moses (1962) suggested that firms located outside the center could thus offer their workers a lower wage, which they would be willing to accept, due to their decreased commuting cost [20]. DiPasquale and Wheaton expand that notion into a theoretical system of decentralized employment centers where firms follow that pattern [11]. Such locational advantages, however, do not go unnoticed by the land market, and rents rise until a new equilibrium of (lower) combined rent and transportation costs are reached. In pursuit of labor savings, firms bid up the price of such desirable, decentralized locations. The research in this thesis tests this theory of the relationship between location and labor by testing the correlation of labor access and office rent.

The competing forces of agglomeration and decentralization lead to a polycentric urban environment, typically with a dominant core and many smaller surrounding employment subcenters [14]. Subcenters are only relevant in this thesis insofar as they represent unique locations that influence the rents of offices within them. McDonald (1995) defines such subcenters as locations that exceed a certain employment density threshold [16]. Other research then uses such subcenter information in office rent models [10, 18]. The research presented in this thesis disagrees with the methodology of attributing rent to a locational dummy variable indicating whether a building is in a subcenter or not. The use of locational dummies without some theoretical justification absorbs the variation in locational rent without describing a substantive phenomenon. Alternatively, this thesis proposes the explicit and descriptive quantification of locational rent with respect to the theories of agglomeration and decentralization. Locational variables are constructed to reflect those phenomena and tested through the use of a spatial hedonic value analysis.

The model put forth here builds upon the literature of hedonic property value analysis, using the metropolitan Boston market for office space as a case study. Hedonic analyses are those which treat the price of a good as the sum of the prices of its constituent parts. By regressing the price of the good, in this case office space rent, on its constituent parts - building and neighborhood attributes - the implicit prices of those attributes are so derived [26]. While this technique is well developed with respect to such physical characteristics, it is less so with respect to the other well know determinant of property value, location.

In the literatures on hedonic modeling of both housing and office markets, the treatment of space is still unsystematic and unrefined. The locational variables cannot be directly compared across studies due to their employment of different spatial units. A unique contribution made by this research is in its treatment of location. Location is broken down into three main components of characterization: area attributes, distance measures, and spatial structure. With respect to the use of area information, there are many units by which data such as income or employment are reported. Waddel, Bailey, and Hoch use the census tract summary level of neighborhood data to attribute to residential property values [35]. Other studies, such as Wheaton and Torto (1994), Clapp, Pollakowski, and Lynford (1992) rely on different analysis *zones*, which are often transportation analysis zones of local planning jurisdictions [10, 11]. Yet another common level of data aggregation used in hedonic models is the United States Postal Service's ZIP code boundary [13]. In all of these cases the neighborhood characteristics are found to be significant contributors to the price model, yet are problematic. There are countless different areal units of data aggregation. First, they are locationally specific and thus difficult to compare research based on data in different places. For example, the demographic makeup of census tracts in Houston are not of equal size shape nor character as those in Denver. Secondly, they change over time; the US Census adjusts their boundaries, as does the postal service. Thirdly and most important, these arbitrary units of space both in terms of size and data aggregation have little if anything to do with any underlying economic or spatial phenomena. Distance units, however, arguably have more to do with the phenomena of interest and are more consistent between studies and across space. The most common measure is distance to the CBD, which is included in most spatial models [8, 10, 13, 27, 29, 35], but newer research extends the assumption about the productivity of the central place to include other nodes and geographic features. Waddell, Berry and Hoch (1993) use distances measures relating to airports, freeways, hospitals, and retail in their hedonic model of residential property values [35]. Studies of office markets such as Sivitanidou (1996) find access to other subcenters to be a significant explanatory variable in rent [27]. The research in the field is beginning to incorporate richer descriptors of location, yet none of the above explicitly model the underlying spatial structure in the data [7]. Spatial structure in data is a statistical artifact of a complex spatial process such as a neighborhood decay or growth on an urban boundary

[4]. A few researchers have begun to incorporate sophisticated techniques in spatial statistics to address spatial dependence. The statistical phenomenon that results from the effect of a spatial process is spatial autocorrelation [2, 8]. It is the condition that exists when the covariance of the error terms of a regression are not zero (as is assumed to be the case in regression analysis), but is a function of the spatial proximity among observations. Can and Megbolugbe (1997) find this to be the case in their estimation of house price indices [8]. Controlling for spatial autocorrelation in the data eliminates bias and improves the efficiency of the regression coefficients.

The research put forth in this thesis builds on these models by presenting a new methodology for capturing the value of location. It builds upon previous research in hedonic modeling using geographic information systems (GIS). Waddell, Berry, and Hoch (1993) and Sivitanidou (1996} begin to incorporate GIS as an analytical tool by using them for precise line of sight and road network distance measures [28, 35]. This thesis goes a step further and proposes a shift in the thinking about which variables at what spatial scales appropriately describe location. The neighborhood units of this thesis avoid these problems by first using as disaggregate source data as possible, and then aggregating the data into units constructed with regard to testable theory. This methodology allows the effect of a given neighborhood characteristic at a certain *spatial resolution*, to be estimated and shown to be significant or not. Furthermore, the construction of neighborhood variables calculated at specific radii in a grid cell model is repeatable across political jurisdictions and through time.

CHAPTER 3:
DATA & METHODOLOGY

MAJOR FINDINGS

The major results of the combined analysis confirm the initial assumptions of the thesis -that the locational attributes of an office building contribute to its rent. The explicit quantification of location explained an additional ten-percent of the variation in the average rent for office space. Locational factors found to be significant were access to highways, labor, shopping centers, and other offices. Separating the data into two geographical subsets {inner region, suburbs} further refined the spatial finesse and added to the strength of the overall model. Use of GIS then permitted the visualization of the findings as a map of potential rent spanning the entire study region of metropolitan Boston. The map dramatically reveals the capacity of the model's sensitivity to space, predicting known neighborhood rent spikes such as Back Bay, and suggests others that may not be as well known. Based on the parameterization of locational factors the image below projects a surface of potential office building rent. A detailed explanation of the process by which this was reached follows in the following section.

Data Used in the Model

To arrive at a more precise model of locational rent than is offered by previous research, a case study of the office market of metropolitan Boston was used. The following chapter discusses the data used in the process of developing the model building process as well as those variables adopted into the final model. For variable definitions please refer to the data dictionary in Appendix A.

The variables break down into three main categories: building characteristics, demographics, and geographical data. The building characteristics were derived from a dataset of 2269 buildings [TWR], 989 of which fall within the greater Boston study area, and have at least one valid rent observation. The fields included in the data are build date, renovation date, floors, net rentable area, class, and exact location given in latitude and longitude. Age of the building since its build or renovation date is suggested as a control variable, and is found to be significant. The number of floors of a building is meant to serve as a prestige characteristic. Net rentable area (NRA) is postulated to be valuable to firms looking for large blocks of contiguous space [36], and its square root is taken to compress the distribution. The effect is to place more weight on smaller differences in building size. The data also includes a class variable; a dummy for Class A is constructed. Building class is considered an absolute and independent adjustment, raising the value of the space without affecting the contribution to the rent of the other building characteristics.

The demographic data included data on median household income, population, and employment statistics, both collected at the census block group summary level. This is typical of the demographic data used in similar research [29, 30]. The reason for using the income data is to test for affluence as a prestige amenity. Little theory exists on why income is a determinant of locational demand other than that the chief operating officer (CEO), or other decision maker desires a location near his/her own house, which is likely to be in a high income area. Worker data is used to gain an estimate of the relevant labor force to test the theories related to commuting mentioned in Chapter 2. Workers are defined as the sum of five census employment categories identified as those that are likely to work in offices. They are generally categories including executives, management, and other professionals. For a complete listing, please see the data dictionary in Appendix A.

The other data category used to calculate additional locational variables is geographical information including network and event data. Transportation infrastructure data includes road network data was made available from the U. S. Census Bureau's Topologically Integrated Geographic Encoding and Referencing (TIGER) division [34], and subway stop locations from the Central Transportation Planning Staff (CTPS) of the Boston Metropolitan Planning Organization [9]. The retail dataset came from the National Research Bureau (NRB) [23] and includes the location and size of 610 shopping centers in the Boston metropolitan area. The reformulation of these data into more spatially illustrative variables is discussed at length in the methodology section.

Proposed Model

The proposed model, which characterizes locational rent, builds on the generic form of a hedonic model shown below.

$$Rent = \alpha + \beta_1[x_1] + \beta_2[x_2] + \varepsilon \quad (1)$$

In this generic form of the hedonic model for office rent, rent is equal to the sum of the matrix of building attributes $[x_1]$ and the matrix of spatial attributes $[x_2]$ within a specified distance. The building characteristics include age, which is the lesser of the time since the year built or the last renovation, height in floors, the (square root) of the net rentable area, and the class [A or B/C] of the buildings. These are used as control variables for the office space in each of the 989 eligible buildings such that the quality per unit is constant and thereafter treated as though it were a commodity. The matrix of spatial attributes includes the census demographics and the spatial relationships between the office buildings, shopping centers, and access measures to transportation infrastructure.

This model is expected to explain the variation in office space rent more accurately than traditional spatially inarticulate models and in doing so derive implicit prices for locational attributes. The derived prices can then be used to predict locational rent across the entire study area. The following chapter describes in detail how the techniques of hedonic modeling combined with GIS are used to test the thesis hypotheses and predict a surface of potential rent.

METHODOLOGY

The methodology employed in determining the locational rent of office space in metropolitan Boston was twofold; it incorporated economic and spatial modeling. A type of econometric modeling, hedonic price modeling, was augmented with the use of geographic information systems for data analysis. In hedonic modeling, the price of an object is broken down into the prices of its constituent parts. This analysis combines such econometric modeling with a tool for spatial analysis, GIS, to arrive at a hedonic model of office rent that explicitly incorporates location.

Econometric Modeling

The basic technique used in the hedonic model was Ordinary Least Squares (OLS) Regression. Simple linear regression was selected because there was little theoretical reasoning to suggest otherwise, and it followed the hedonic literature [7, 8, 28, 29]. Additionally, linear OLS has the convenience of having the same units as the original data, enabling a straightforward interpretation of the relevant relationships between the coefficients and the scale of the dependant variable, dollars. An alternative would have been the use of a log-linear model, but preliminary results indicated that it produced weaker results. The creation of a hedonic model of office rent involved a number of steps. The first of which was to control for buildings' physical characteristics. Once the variation in the rent for office space due to its size, age, class, and height were isolated and priced, a constant quality unit was created for further comparison. The difference then between a building's rent and that portion for which was so controlled is assumed to be due to location. From this point, location was broken down into constituent parts: access to labor, transportation, retail, other firms, and neighborhood income. The result of such work was to derive and explain the implicit price of the locational component of an office building's rent.

Two Models

While spatial variables of different radii were created with the understanding that some were more meaningful in the city and others in the suburbs, it became clear that modeling the data as one did not reflect the reality of spatial processes. One critical aspect of the model, is the division of the dataset into two categories; buildings are either in an inner region or a suburban one. The following equations describes rents for both the inner and suburban regions as,

$$Rent_{I,S} = \alpha + \beta_1[x_1] + (d\beta_{2,I} + (1-d)\,\beta_{2,S})\,[(d)x_{2,I} + (1-d)x_{2,S}] + \varepsilon \qquad (2)$$

where d is a dummy variable and is 1 for inner buildings and 0 for suburban buildings. The constant in the equation, α, and the structural characteristics, $[x_1]$, are constant between the regions. However, $[x_2]$ refers to spatial variables relevant to each zone. The inner region is defined as Boston, Brookline, Cambridge, and Somerville, and the suburbs as the remainder. The data was separated this way because there is a break in the distribution of buildings along these lines, and these four inner cities are markedly different from the rest of the region. The empirical results of the separation affirmed this categorization of the metropolitan area.

Each model was run separately, and it was discovered that very different resolutions of the spatial variables created the best fit. Although space and distance were found considerably different in the city than in the suburbs, the data still represented one regional market for office space. The two models were thus folded back into one, but to keep the submarket effect in the combined model, each spatial variable was interacted with a dummy [1=inner, 0=outer] for its location as inner region or suburban, while the structural elements were not. It was thus possible to apply variables of different spatial resolutions only to *where* they were most appropriate. This is discussed at greater length in the following chapter on results.

Time

Another econometric method used in the study was the modeling of rents through time. This was done for two main reasons. One is that nine years of data provide many more observations. The other reason was to approach the question of whether or not the model, and thus locational value, is stable over time. The office building dataset contains nine years worth of rent data, but unfortunately, it is not complete. Therefore, observations were appended for each year that a building had a valid rent observation, and including a dummy variable for that year. This was modeled as

$$Rent_{I,S,,Year} = \alpha + \beta_1[x_1] + (d\beta_{2,I} + (1-d)\,\beta_{2,S})\,[(d)x_{2,I} + (1-d)x_{2,S}] + \beta_{3\text{-}11}[x_3] + \varepsilon \tag{3}$$

where $[x_1]$ represents the vector of building structure characteristics, $[x_2]$ that of the inner and suburban spatial variables, and $[x_3]$ that of the vector of annual dummy variables. In this way, it was possible to use all available rent observations, and capture the effect in aggregate of each year as an independent variable. To test the stability of the significance of the spatial variables, each year was modeled separately. Then when all years are modeled together, the coefficient on the annual dummy is interpretable as a general indicator of level, and covers the entire metropolitan real estate market. The annual dummy is presumed to shift the intercept of the equation, but not the value of the coefficients.

GIS

The other methodological component of this thesis was the use of a geographic information system(GIS). A GIS is an integrated system for the collection, storage, processing, and display of spatially referenced data. It can be thought of as a digital marriage between a database and a map, allowing the user to interact with data either graphically or tabularly. Much of its benefit as a tool for analysis and research comes from the use of a map as a graphical metaphor which permits varying sources of data to be related *spatially*. This is accomplished as data that are spatially referenced are overlaid on top of one another as they are in space. The orthogonal view reveals by location as opposed to by observation. This facilitates the simultaneous observation of data values of different sources. Furthermore, the data are not just superimposed images; the system maintains the co-location of events in separate (data) layers in its memory, which allows the user to make calculations across datasets conscious not of record number, but of location. This is analogous to relating or joining two data tables by matching up records via like values in a common field, the common field in this case being location.

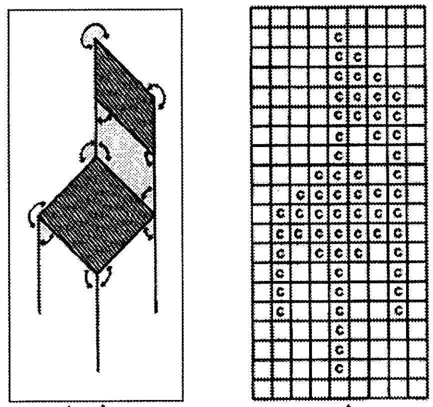
There are two different types of GIS, vector and raster, both of which were used in this thesis. In a vector GIS, data is stored topologically. The elemental data unit in a vector GIS is the point, out of which lines and polygons are constructed. Every element a user sees in the map is comprised of these basic geometric elements, and the data is correspondingly associated with its relevant topological counterpart (point, line, or polygon). This type of GIS is particularly well suited to the analysis of networks such as roads, and is therefore the dominant GIS used in transportation analysis. The other type of GIS, raster-based, uses completely different means of characterizing space. In a raster GIS, space is divided up into a grid of cells like a sheet of graph paper. Instead of a data layer consisting of topological units and their associated value, it is a sheet of cells each of which contains an individual value. GIS analysis was employed in the model for its ability to maintain the richness of discrete locational information along with the actual data values.

There are several ways of accomplishing the characterization of location for the purposes of building a hedonic model. Traditionally, the means of characterizing location rely on previously defined definitions of space, such as town boundaries or census geographies. However, such boundaries do not necessarily have anything to do with any spatial process, such as the expansion of the financial center in downtown Boston. That process could be postulated to be sensitive to distance from the capitol, or the airport [10,16,28], but certainly not zip code or census tract boundaries. Furthermore, such forced aggregation into arbitrary geographies obscures variation within and artificially magnifies variation across their boundaries. For example, take the case of the population density of the state of New York. When addressed as a whole, the number is much lower than what might be considered typical for New York. Viewed at a county level, differences begin to materialize between counties, relaying a more accurate measure of the density people actually experience. However, there is still much variation in density within the county borders, and a more gradual change between them that are obscured by the county average. A picture of density aggregated by city further resolves this issue, revealing the variation inside the county and mitigating the sharp differentials at the borders. To mitigate these distortions, a GIS was used to analyze *all* locations across the study space, via the grid cell model, at a resolution much smaller than the smallest standard unit if area (census block group). The subsequent paragraphs narrate the modeling as an interactive process of GIS data modeling and econometric modeling.

GIS Analysis

The first step of this iterative process was to define a study area, and establish grid parameters. The area chosen was 2000 square miles bounded roughly by a rectangle that includes the outermost radial highway in Boston, Interstate 495. For the purpose of the raster-based analysis, this area was then divided into 500,000 grid cells of 100 meters a side. There is a trade off in the selection of a cell size between resolution and performance. The study area was defined such that the cells would be small compared to other more aggregate polygons such as towns or block groups, but large enough to keep the dataset a manageable size. The choice of 100 meters per side meant that few cells would contain more than one building, so as not to lose the benefits of using disaggregate data. It also meant that most features of the landscape would be preserved, such as the Charles River and the Fort Point Channel. A smaller size would have yielded even more resolution and thus accuracy, but at below 10,000 square meters of area per cell, the calculations would have required additional (unavailable) computing resources.

After the selection of the grid parameters, the relevant data coverages were converted from their original vector format to a raster format. The GIS simply attributed each underlying cell with the data from the associated polygon, as is depicted in the figure below.



vector image rastor

P.A. Burrough

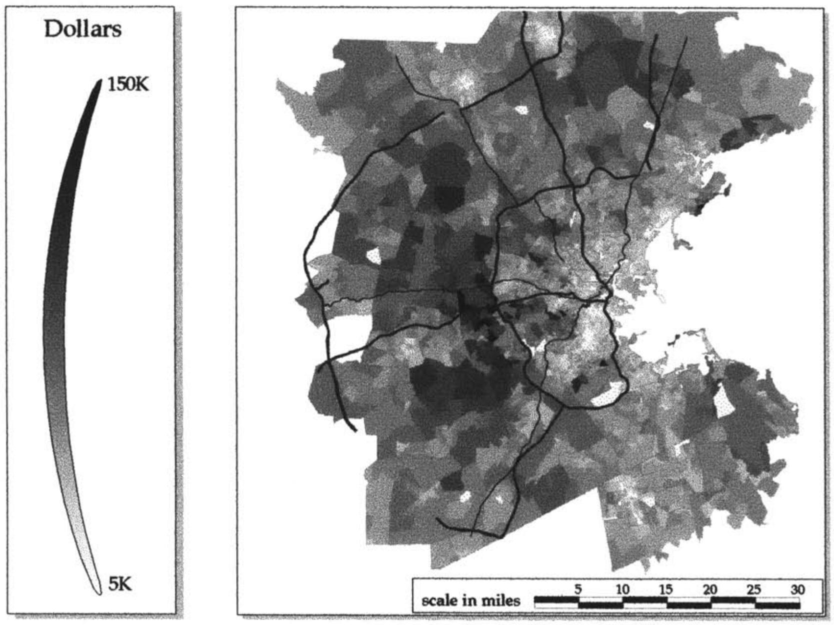
Figure 2.

Demographic Variable Construction

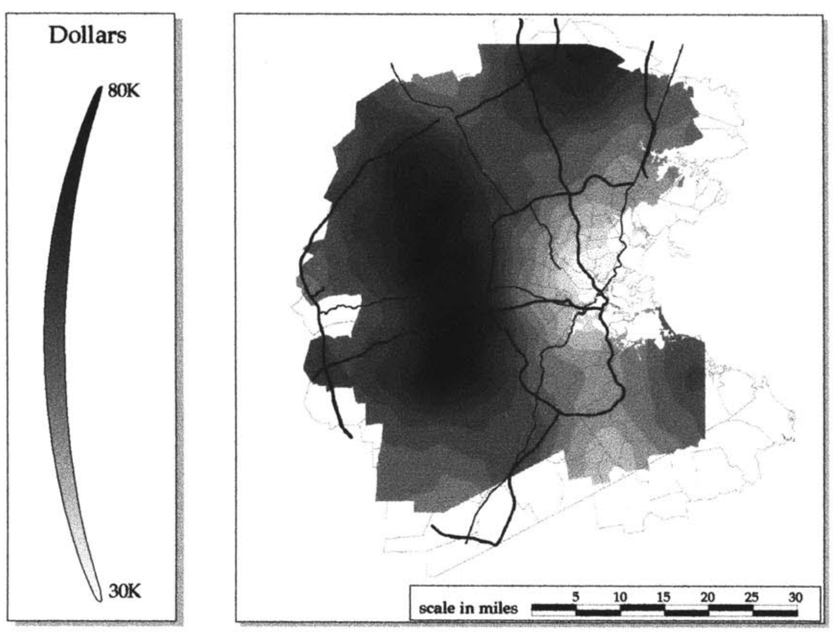
As was mentioned in Chapter Two, there are three main sources of data in this analysis: demographic data, office building and shopping center locations, and transportation infrastructure. Each of these datasets was manipulated with the GIS to produce a more accurate depiction of the location it describes. The first datasets converted were the census-based demographic information at the block group summary level, the finest level of resolution publicly available. The variables chosen were median household income, and 'workers', where workers are defined as the sum of persons employed in those industries where an office building is the likely workplace. Median household income was used as a proxy for the general prestige of a neighborhood, and workers a measure of labor supply. Once rasterized, the values reported by the census per block group polygon were re-aggregated into neighborhoods of different spatial resolutions. The product of such data manipulation was a continuous surface of a given variable. This is a more realistic portrayal of location, as neighborhood phenomena exert their influence across space, often without regard to any specific boundary, especially non-political ones.

The following figures are included to demonstrate the difference between the original census geographies and those proposed in this thesis. Figure 4 portrays how the picture of the distribution of income is changed when the original data (Figure 3) is rasterized and re-aggregated at a radius of a ¼ mile. This is the result of assigning every cell within a given block group the block group value and then recalculating that cell value as the average of all of the cell values within the 5 mile radius.

Median Household Income



Vector
Census Block Group
Figure 3.



Raster
Five Mile Average
Figure 4.

Source: US Census

Figures 3, 4 show a dramatic difference in descriptive ability made by the recalculations of the census variables. The purpose of the income variable is to gain a picture of the overall attractiveness of an area. The fact that certain small high income neighborhoods are completely washed out, and that the affluent suburban areas are exaggerated is exactly the point. Thus, although the ¼ and 5 mile averages are ostensibly measures of the same phenomena, as the regression results in Chapter 4 show, the different spatial configurations of these variables yield quite different results.

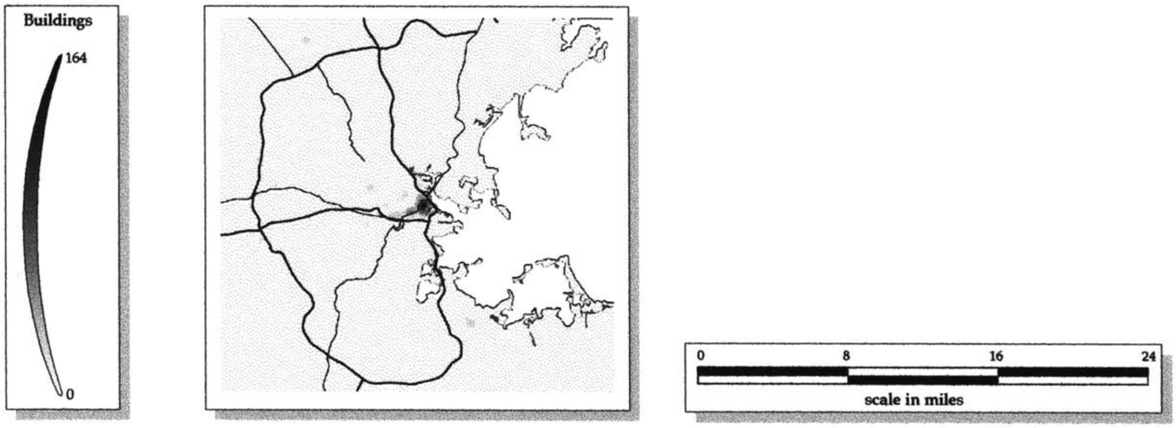
The other variable originally defined as is a census polygon is workers. Workers were re-aggregated somewhat differently, since workers are a count variable rather than a descriptive statistic of a distribution. To calculate the number of workers within a radius, the block group data was converted into a density measure. Each cell was assigned the number of workers per unit area as opposed to the simple block group value. Summing the cell values across different radii yielded a count within the specified radius.

Event Data

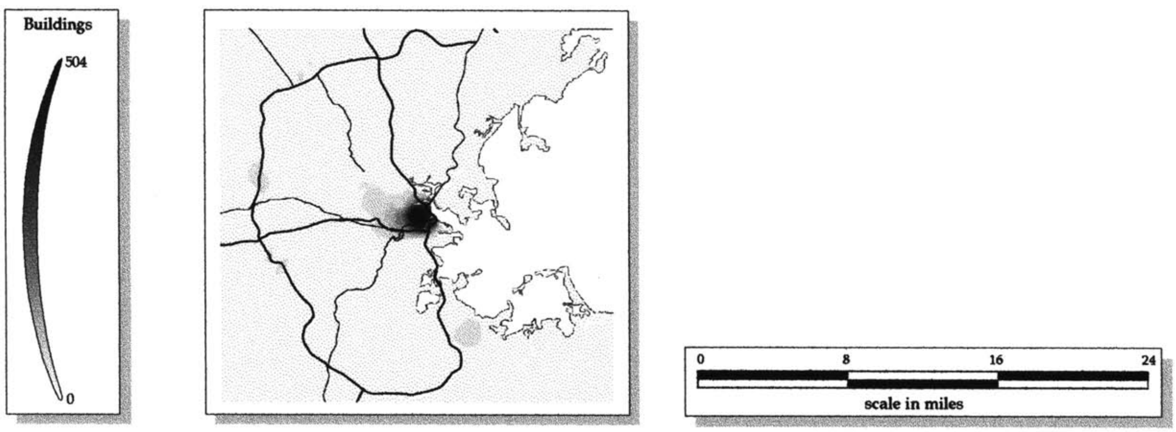
The second group of spatial variables to compute were those coming from point data. Office building locations and shopping center locations can be thought of as *events* in that for every location (cell) across the study space, a building either exists or it does not. Creating the spatial variable for event data consisted of establishing for each cell the number of such events within a specified distance.

Part of the utility of GIS is the ability to create multiple measures of the same variable, but along with that power comes the difficult decision of which radii to use. This analysis was more explorative in this regard than definitive. Figures 5-7 show the number of office buildings within a few radii, chosen for each variable according to the reasoning explained in the following paragraphs.

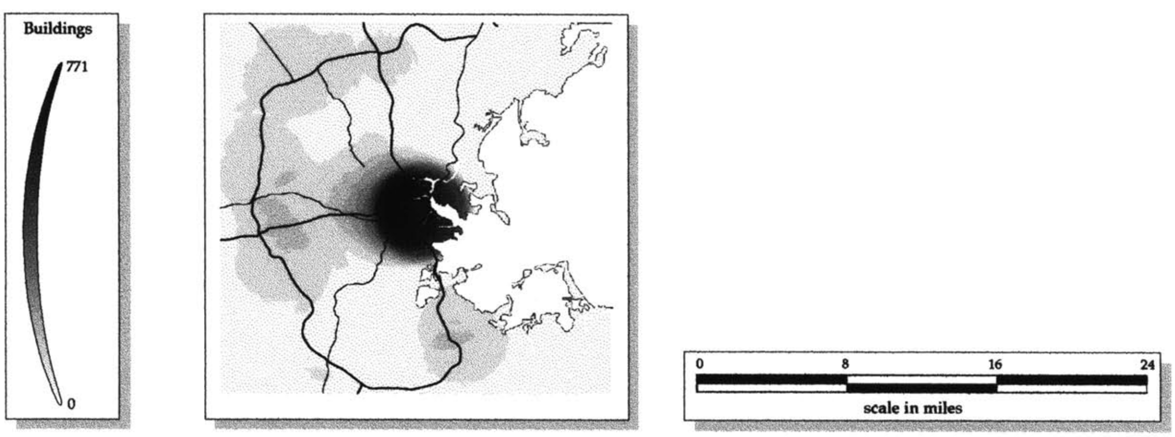
Office Buildings



Office Buildings Within .25 Mile
Figure 5.



Office Buildings Within 1 Mile
Figure 6.



Office Buildings Within 3 Miles
Figure 7.

Source: Torto Wheaton Research

Worth noting at this point is that neither the simple smoothing nor aggregative measures described above are postulated to be optimal. Additional work on appropriate smoothing and aggregative algorithms such a distance decay function, for example, might better reflect the underlying spatial process. What is important for this thesis is that such measures can be captured under this methodological framework. Despite its simplicity, these methods worked reasonably well in the final model, which is discussed further in Chapter 4.

Variables & Spatial Resolution

It was assumed that measures of smaller radii, or finer resolution, were necessary to capture spatial phenomenon in the inner region, while much larger radii are more appropriate in the suburbs. This assumption was based on the difference in overall character of the two places. In the inner region, density is much greater and a smaller share of transportation is by automobile. The opposite is true in the suburbs, where most activities necessitate the use of a car. Therefore, an event or process may exert influence over a small area in the inner region, and a large area in the suburbs. Furthermore, for some of the variables there was more theoretical justification for finer resolution than others.

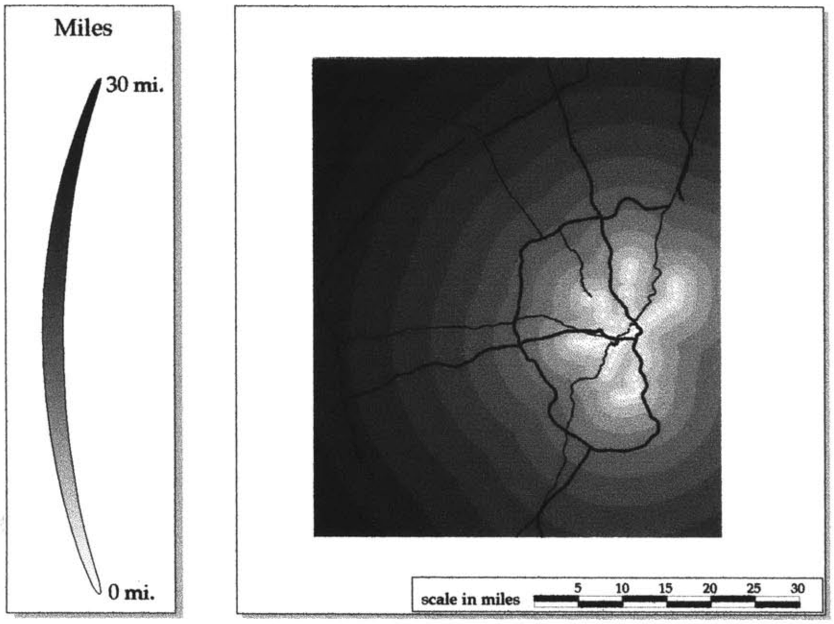
The relevant distances to use for the measure of labor supply come from a basic analysis of general commuting patterns in Boston. Since it is a dense city with significant congestion, the actual distances are not great. Three miles was chosen as the effective commutable distance within the central areas of Boston for those commuting to the center, while five and eight miles were considered reasonable commuting distances by car in the suburbs. These numbers were roughly estimated by observing from the census that the average reported commuting time was 20 minutes. The translation of time into distance, however, depends very much on location. There are two reasons time is locationally sensitive. One has to do simply with mode choice; a large number of people in the inner region use mass transit in their commute. They travel a much shorter distance than those commuting by car in the suburbs for a given amount of time do. The other reason is that even by car one cannot travel nearly as fast in the inner region than in the suburbs. This is indicative of the level of congestion present in the older, central cities in Boston. In the same twenty minutes, a suburban commuter might travel approximately seven to ten miles, yet an inner city resident may traverse only three or four miles. Therefore, different spatial measures were applied.

With respect to the spatially continuous variable of income, there was little theoretical guidance, other than the general tendency for increased heterogeneity in the inner cities. Thus spatial variables were assumed to be significant at larger scales in the lower density suburbs than in the city. The radii chosen for the event information, however, were much more informed. The purpose of the neighborhood measure of office buildings was to capture any effects of agglomeration economies. For central locations, especially in Boston, this means walking distance, which means that such effects will be captured using very high resolutions. Developers and planners consider ¼ mile the threshold distance after which people are no longer likely to walk. Therefore, office building neighborhoods of ¼ and ½ mile radii were created for the inner cities. In the suburbs, however, walking is much less likely and once one commits to drive, the sensitivity to distance falls sharply and much larger radii are appropriate for testing. It should be noted here that despite its intention as such, this variable as constructed is not necessarily a measure of agglomeration economies. Simple economies of urbanization or specialization may simultaneously affect rent, an effect that is not differentiated in this research. Agglomeration economies referred to in the context of this research may consist of any or all of these contemporaneous effects. Similarly, the effects of nearby shopping centers were postulated to abide by the same generalization and subject to the same caveat regarding confounded effects.

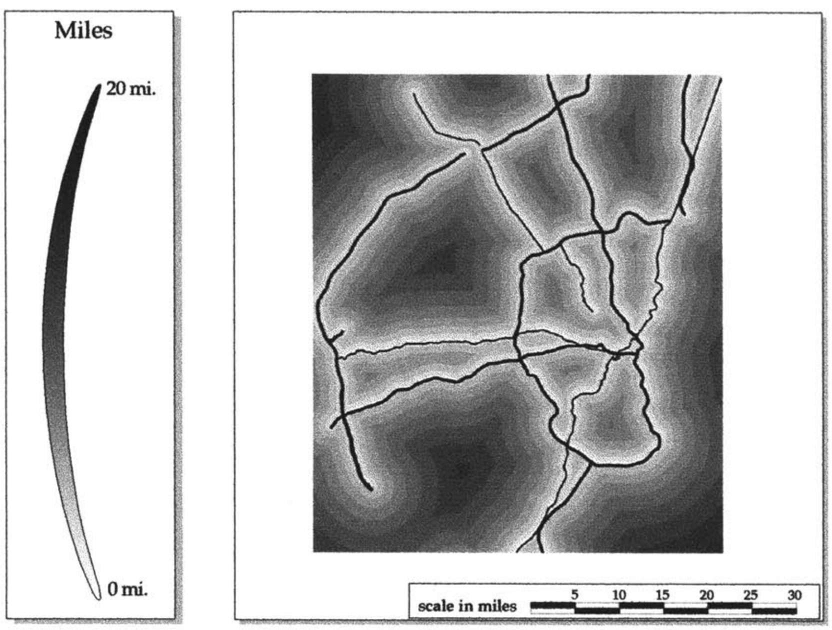
Distance Variables

Another set of spatial variables made possible using GIS are distance measures. It is often postulated that a large portion of the value of a given place is attributable to proximity to the local transportation infrastructure. To test this intuition, the distance from each cell to the closest highway and (separately) subway stop was calculated. Although distance measures were created for both, it was assumed that highway access is appropriate for the suburbs and subway access for the central city. Another measure significant to traditional urban economics, and thus calculated, is distance to the central business district (CBD). One of the more elegant features of calculating such distance measures is that no arbitrary radius need be chosen. However, raw distance may not be as fitting as some function of distance. Thus, for the purpose of this study, square and square root of the straight-line distances were calculated. The square root of distance better captures the relevant behavior in that there is much more sensitivity to the first few miles, after which it declines. This is to say a commuter is very sensitive to being within one to four miles of a freeway, but beyond that, the difference to the commuter between five and fifteen miles is minor. Figures 8 and 9 depict these distance buffers.

Distance Measures



Distance From Subways
Figure 8.



Distance From Highways
Figure 9.

Source: CTPS

To complete the iteration back to the statistical modeling, it was necessary to collapse all of the spatial variables into the building dataset. The GIS enabled the attribution of the value of each derived variable in each cell to each building located within that cell.

Attribution of Grid Data to Office Building

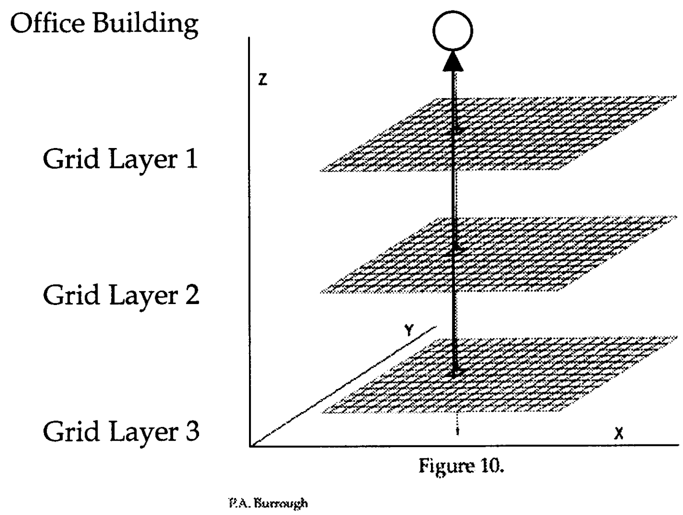


Figure 10.

This yielded a building dataset which contained the spatial attributes as well as the building attributes for the area in which each building was located, thus explicitly defining the location of the structure. Regression analysis of the relationships between the spatial variables and building characteristics were then performed as an interactive process. A spatial variable was kept in the model if it was statistically significant from zero at an alpha level of 0.01. The model grew and changed with the addition of each spatial variable, until a final model was reached.

CHAPTER 4
RESULTS & INTERPRETATIONS

Chapter Introduction

Each of the three main objectives of the thesis is addressed in this chapter of results. The modeling efforts yielded a refined hedonic model of locational rent that quantifies urban space. The variables used to construct the model were successful in bearing evidence to support answering each of the original research questions in the affirmative. Ultimately, the product of the research is the map of potential office space rents shown below. Potential is the expected rent for an office building if one were located at *any* particular place. This is different from plainly predicting office rents for existing buildings because it is a prediction for *all* locations, whether or not office buildings or any commercial activity exists or not.

Surface of Potential Rent

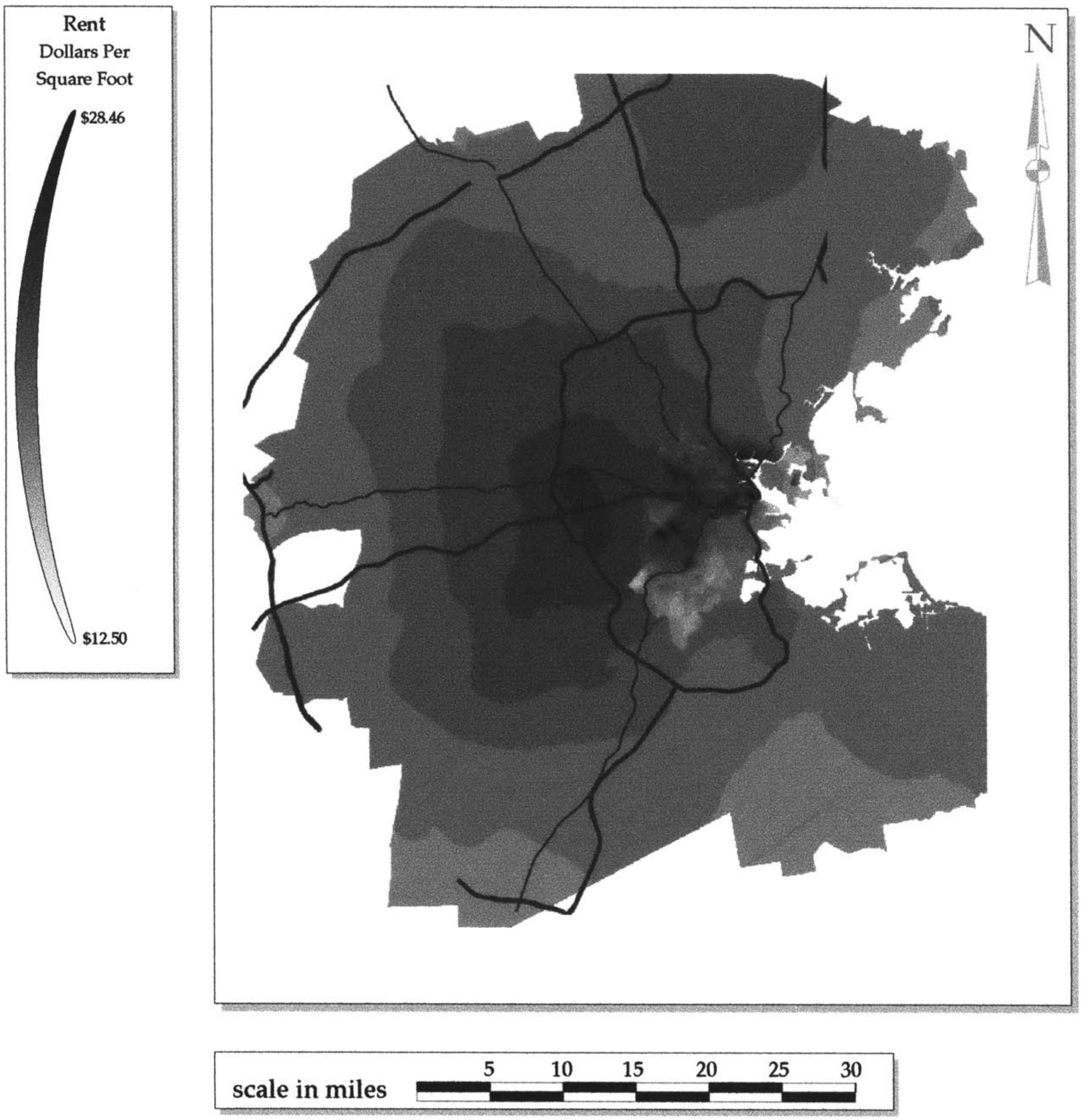


Figure 11.

The results presented by the above map were reached through a several stage process generally described in the previous chapter. However, in this chapter, some of the interim results of significance are discussed, beginning with the base model of structural characteristics only. That is followed by a brief discussion of the regression output that led to the splitting of the dataset into two models. Then the final spatial model is presented. After the presentation of the final spatial model, each of the original research questions is discussed with regard to the regression output of the modeling process. The chapter concludes with a discussion of the map of potential rent printed above.

Base Model: Structural Characteristics

The results of the modeling process are detailed below, beginning with a non-spatial model in which all of the buildings were used for all nine years. The coefficients are given (β) along with their relative significance (t), an absolute value of which over 1.8 is considered to be statistically significant. To give an indication of a typical value, column four shows the coefficient multiplied by the standard deviation of the distribution of the original variable.

$$\text{RENT} = \alpha - \beta_1[\text{Age}] + \beta_2[\text{Class\_A}] + \beta_3[\text{Srrtnra}] + \beta_4[\text{Floors}] + \beta_{4\text{-}11}[\text{YearDummy}] + \varepsilon \qquad (4)$$

Table 1: Base Model

Adjusted R Square	.39797
Standard Error	4.55623

Variable*	β	t	β * σ	Average in Dataset
AGE	-.028931	-8.073	$-0.50	21 Years
CLASS_A	3.192946	17.008	N/A	N/A
FLOORS	.305538	21.458	$ 2.05	6 Floors
SRRTNRA	.002947	4.526	$ 0.50	101,550

*Please refer to Data Dictionary in Appendix A for complete description of variables.

The base model reported above includes only the physical attributes of each office building. The relative magnitude of the variable Floors compared with Size suggests a mis-specification error related to a downtown effect. The buildings downtown exhibit higher rents than would otherwise predicted, especially in a non-spatial model, since there is definitely something very particular and valuable to a downtown location. It also happens to be where buildings are the tallest, which is why their otherwise inexplicable rent is attributed to their height. This effect is stronger than total size in square feet, because there are enough large buildings in the suburbs to mitigate the potential size effect, while there are almost no buildings in the suburbs that could be called skyscrapers. It is difficult to realistically believe this model's prediction that the rent is more a product of a building's height than its total square footage. An approximate $3 premium for Class A space is, however, quite reasonable, as is the modest decay in value with building age.

Spatial Models

As space is added into the modeling process, the regressors become more intuitive. While the changes in the total R^2 are not great, the magnitude and influence of the individual variables is more in tune with reality. Part of the iterative process of reaching a final model was the experimentation of explanatory variables of differing resolutions. In the table below, where the β values are reported with their corresponding t statistic in parenthesis, intermediate results are shown. It became clear that smaller resolutions were appropriate for the buildings in the inner region, yet forcing those same resolutions on the suburban regions as well did not make sense. Model 1 below shows just that, and similarly Model 2 forces the larger resolutions constructed for the suburban region on entire dataset. The conclusion was therefore to split the dataset into two parts, and apply to each region the appropriate variables (Model 3).

TABLE 2: MODIFICATION OF RESOLUTIONS

	Base Model Structural Model	Model 1 Inner Variables	Model 2 Suburban Variables	Model 3 Final Model (Split)
(Constant)	16.00 (72.3)	9.28 (10.2)	9.06 (17.2)	7.00 (15.3)
AGE	-0.0289 (-11.3)	-0.036 (-10.6)	-0.037 (-11.0)	-0.0365 (-11.2)
CLASS_A	3.19 (17.0)	3.14 (17.9)	3.21 (18.0)	3.03 (17.7)
FLOORS	0.305 (21.5)	0.088 (5.4)	0.19 (12.32)	0.084 (5.26)
SRRTNRA	0.00295 (4.526)	0.007 (11.2)	0.0047 (7.4)	0.0075 (12.0)
IINC_25M		7.36e-5 (21.0)		7.90e-5 (17.6)
INRB_50M		0.159 (3.3)		0.239 (4.00)
IOFF_25M		0.017 (9.3)		0.189 (8.34)
IWRK_3MI		1.94e-5 (15.9)		2.95e-5 (7.57)
ID2HWY_5		-0.024 (-8.1)		-0.033 (-3.81)
INNERDUM				2.25 (2.81)
OINC_5MI			8.55e-5 (12.4)	9.99e-5 (14.1)
OOFF_3MI			0.0035 (7.8)	0.012 (4.49)
OWRK_8MI			6.10e-6 (10.9)	5.55 (7.84)
N	5865	5743	5865	5865
Adj r2	0.398	0.49	0.46	0.513
F-value	324	325	315	295

Small adjustments in the coefficients, as those observed between models above, may not change the whole model vary drastically, but they are significant locational changes, which in effect alter the predictive surface. In terms of differences between the base and final models, the change in the coefficients of the size related building attributes is immediately recognizable. The premium on floors drops to a more reasonable $0.08 from $0.31 in additional rent per floor. This means that the spatial model only places a substantial premium on the few truly tall buildings in metropolitan Boston. Since it is difficult to postulate any productivity gains resulting from height, this result is interpretable as a prestige factor which really only applies to the tallest buildings. The reverse scenario, however, applies to the case of the other building size measure, net rentable area. Therefore, the finding that the coefficient on size in the spatial model attributes a larger portion of the rent to the total size of a building makes sense. It is reasonable to assume that larger blocks of contiguous space are valuable to firms, especially large firms, so that they can co-locate an entire company or department.

The final spatial model represents a significant improvement over the model composed of only the building characteristics only, explaining approximately 10% more of the variation in office space rent. However, what is most important to this research is not necessarily the goodness of fit of the model (R Square), but the significance of the variables. The included variables in the model are interpreted in the following section. In equation (5), the I_ prefix denotes a variable only applied to the inner set of buildings, while the O_ prefix denotes those in the suburbs. The annual dummy variables' coefficients in Table 4 give the annual shift in the general market rents relative to 1996.

$$\begin{aligned} RENT = \ & \alpha - \beta_1[Age] + \beta_2[Class\_A] + \beta_3[Size] + \beta_4[Floors] + \\ & \beta_{14}[Innerdum] + \beta_8[I\_d2hwy\_0.50m] - \beta_9[I\_nrb\_0.50m] + \\ & \beta_6[I\_inc\_0.25mi] + \beta_7[I\_wrk\_3mi] + \beta_{10}[I\_off\_0.25mi] + \\ & \beta_{11}[O\_inc\_8.0mi] + \beta_{12}[O\_wrk\_8m] + \beta_{13}[O\_off\_3mi] + \\ & \beta_{15\text{-}22}[YearDummy] + \varepsilon \end{aligned} \quad (5)$$

TABLE 3: FINAL SPATIAL MODEL

Adjusted R Square	.51313
Standard Error	4.09735

Variable	Beta	t	β * σ	Average Value
Building Characteristics				
Age	-0.04	-11.2	$-0.50	21 Years
Class_A	3.03	18.0	N/A	N/A
Size	0.0075	12.0	$ 0.50	101,500 Square Feet
Floors	0.084	5.2	$ 2.05	6 Floors
Inner Cities Spatial Factors				
Dummy Variable for Inner Cities	2.25	2.8	N/A	N/A
Neighborhood Income: r = ¼mi	7.90e-5	17.6	$ 1.70	$33,000
Other Office Buildings: r = ¼mi	0.02	8.4	$ 0.95	59 Buildings
Number of Workers: r = 3mi	2.95e-5	7.6	$ 2.41	161,400 Workers
Highway Access: Sqrt(distance)	-0.03	-3.8	$ -0.36	500 Meters
Shopping Centers: r = ½mi	0.24	4.0	$ 0.41	2.3 Shopping Centers
Suburban Cities Spatial Factors				
Neighborhood Inomce: r = 5mi	9.99e-5	14.1	$ 3.04	$59,000
Other Office Buildings: r = 3mi	0.012	4.4	$ 0.61	72 Buildings
Number of Workers r = 8mi	5.55e-6	7.8	$ 0.85	237,000 Workers

TABLE 4: YEAR DUMMIES

Variable	B	t	Sig. t
YRDUM88	2.486817	11.111	.0000
YRDUM89	2.450915	11.067	.0000
YRDUM90	1.773288	8.128	.0000
YRDUM91	.504364	2.298	.0216
YRDUM92	-.637902	-2.867	.0042
YRDUM93	-1.904055	-8.541	.0000
YRDUM94	-1.397970	-6.530	.0000
YRDUM95	-.705932	-3.246	.0012

Research Questions

The analysis provides insight into to the research questions presented in Chapter 1. The sign and relative magnitude of the most of the variable coefficients given in Table 3 are consistent with the expectations of traditional urban economics. The model affirms that the spatial demand for space is comprised of the value firms place on the prestige of their location, access to the labor force, economies of urbanization/agglomeration, as well as access to transportation and retail activity for buildings in the inner region. A few of these results, however, are more or less pronounced than was anticipated.

Office rents are correlated with neighborhood income

Income turned out to be more significant a determinant of office rent than anticipated. The difference between the radii, ¼ mile in the inner region and 5 miles in the suburbs, at which the income variable is most significant, is consistent with the model's assumptions. The effect is, however, roughly the same in both regions, contributing a dollar in rent for every $10,000 in neighborhood income. Overall though, the effect is stronger in the suburbs as on average, (especially at a five mile average in this dataset) neighborhood income is much higher. The reason why this is the case is unclear. There are two possible explanations for these results; either neighborhood income extends a degree of prestige to an office firm, or it is evidence for the theory that firms' location decisions are influenced by the residential location of the CEO, which is often in a high income neighborhood.

Office rents are correlated with access to the labor market.

The worker variables nicely confirmed the hypothesis that firms are willing to pay more in rent to locate closer to the labor force. Again, the smaller radius was most effective in the inner region and the larger in the outer region, congruent with expectations.

There is evidence of agglomeration / urbanization economies in the office market.

The last variable common to both subsets of data is proximity to other office buildings. Proximity to office buildings happened to be slightly more powerful an explanatory variable than aggregate square footage of space within a specified radius. A continuous variable of access to space (instead of buildings) fits the theory of agglomeration more closely. This measure may have additional explanations that may not necessarily be related to agglomeration economies. Other effects the variables might be capturing are a general access to urban amenities or general desirability of a location (which may be correlated with the office building variable). It may not be access to other firms that is of value, but some other unspecified factor that influences all firms likewise. Therefore, though a part of the specific advantage of the location cannot be identified, its effect is captured by the office variable in this model.

Office rents are correlated with access to retail opportunities & transportation infrastructure. These last two spatial variables in the model are the least significant, and are relevant only to the inner region. The value of proximity to highways decays with the square root of the straight line distance to the highway itself. As such, it is not a variable without problems as an indicator of commuting costs. Distance from access points might be more relevant. However, it can still be interpreted as such. The fact that it is only significant (at any spatial resolution) for the inner cities was a surprise, as traditional economic theory places considerable weight on transportation access. While still a great asset, it may be that effects of fixed investments of such magnitude are confounded with other variables. In other words, other locational attributes such as office density or retail activity have developed in an area because of the transportation facility, which are captured in the effect of increased rent in this model. Such an explanation, however, is complicated by the difference in access value to the inner and outer regions.

It was anticipated that access to highways would be important for the suburban cities, and access to subways to the inner cities. The latter turned out not to be a significant factor in any of the models tested. Finally, the shopping center data yielded disappointing results compared to the other variables. It was found significant at a small resolution (½ mile) only for the inner region. The lack of strength in the inner region, and its complete absence in the suburbs make it difficult to accept the research hypothesis that firms value access to urban amenity of retail activity -to the extent that it influences rents. Better variables are needed to capture the value known to be related to this phenomenon. While it is easy to discount the effects of access to highways and retail it does contribute to the overall model and differentiate space. This is more apparent when the modeling results are represented graphically via the map.

The model is stable through time.

As was mentioned above, the dataset was created by appending a record for each year for which there was a valid rent observation. This methodology is not as sound as a repeat sales method [3,31], but if the assumption that there is no systematic bias in the selection of observation per year is true, it should give similar results. It also should be noted that only rent was allowed to vary in the model; the spatial data was assumed constant through time. When analyzed separately for each year, however, most of the coefficients are, in fact, stable over time. The table below shows the coefficients from year to year with their corresponding t-statistics in parenthesis.

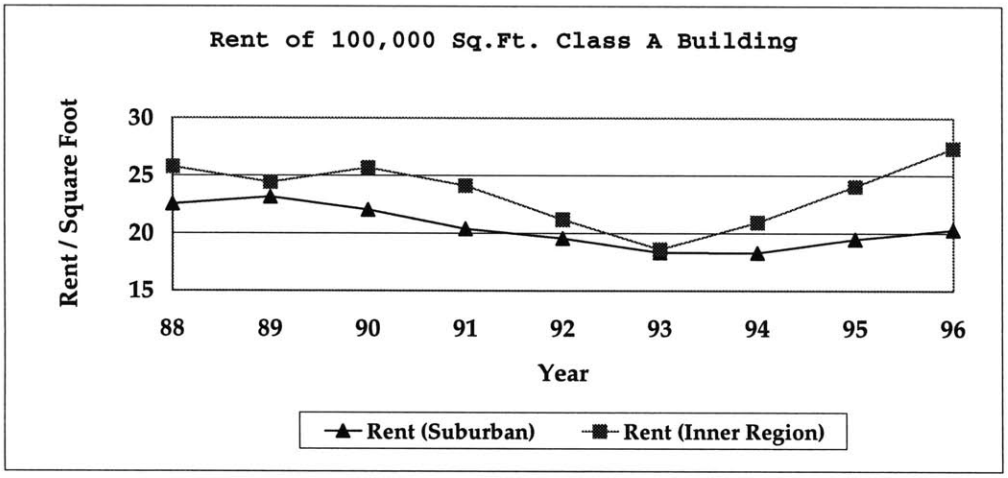
TABLE 4
STABILITY OF COEFFICIENTS THROUGH TIME
β (t)

	88	89	90	91	92	93	94	95	96
(Constant)	8.620268 (6.2)	9.624055 (7.3)	7.954584 (6.1)	6.806538 (5.1)	7.237357 (5.7)	7.73591 (6.6)	6.43932 (4.6)	5.58594 (4.3)	4.99485 (4.1)
AGE	-0.049684 (-4.7)	-0.051763 (-5.2)	-0.053756 (-5.0)	-0.042382 (-4.0)	-0.046874 (-3.6)	-0.0497 (-4.4)	-0.0231 (-2.4)	-0.0217 (-2.6)	-0.03147 (-4.5)
CLASS_A	3.654719 (6.7)	4.180887 (8.1)	3.649221 (7.4)	2.827619 (5.7)	2.668285 (5.5)	2.75859 (6.2)	2.36797 (4.3)	2.42219 (4.6)	2.907792 (6.0)
FLOORS	0.082517 (1.8)	0.1363 (2.9)	0.098895 (2.1)	0.0258 (0.5)	0.0738 (1.6)	0.04121 (0.9)	0.0345 (0.7)	0.1114 (2.2)	0.130342 (2.9)
SRRTNRA	0.007885 (4.2)	0.006412 (3.4)	0.008096 (4.4)	0.010137 (5.3)	0.007991 (4.3)	0.00633 (3.7)	0.00725 (3.8)	0.0065 (3.5)	0.006668 (4.0)
IINC_25M	9.615E-05 (7.5)	0.0001056 (8.4)	8.928E-05 (7.0)	6.261E-05 (4.8)	6.094E-05 (4.6)	6.4E-05 (5.3)	9E-05 (6.0)	7E-05 (4.8)	7.34E-05 (5.6)
INRB_50M	0.206389 (1.1)	0.050715 (0.3)	0.1212 (0.7)	0.126403 (0.7)	0.119486 (0.7)	0.04595 (0.3)	-0.1266 (-0.7)	0.67311 (3.7)	0.81973 (5.3)
IOFF_25M	0.029576 (4.3)	0.03352 (5.1)	0.028119 (4.2)	0.019396 (2.8)	0.018812 (2.7)	0.02003 (3.1)	0.02221 (3.0)	0.00135 (0.2)	0.006778 (1.1)
IWRK_3MI	1.612E-05 (1.3)	3.345E-05 (2.8)	4.298E-05 (3.4)	4.679E-05 (3.7)	2.607E-05 (2.0)	3.8E-05 (3.1)	7.1E-05 (5.9)	3.3E-06 (0.3)	1.29E-05 (1.4)
ID2HWY_5	-0.080175 (-3.0)	-0.07025 (-2.9)	-0.050461 (-1.8)	-0.053162 (-1.9)	-0.000419 (-0.0)	0.00051 (0.02)	-0.0372 (-1.4)	-0.0326 (-1.3)	-0.00856 (-0.4)
INNERDUM	5.459707 (2.2)	1.107061 (0.5)	0.876022 (0.4)	1.2757 (0.5)	2.059143 (0.8)	-0.7396 (-0.3)	-4.7466 (-1.8)	6.84155 (3.0)	5.681094 (2.8)
OINC_5MI	9.95E-05 (4.6)	9.441E-05 (4.6)	0.0001072 (5.2)	0.0001052 (5.1)	9.19E-05 (4.7)	7.8E-05 (4.2)	9.6E-05 (4.4)	0.00012 (6.0)	0.000139 (7.3)
OOFF_3MI	0.014733 (2.6)	0.010328 (1.9)	0.009416 (1.7)	0.006777 (1.2)	0.005688 (1.1)	0.00342 (0.7)	0.00634 (1.1)	0.01112 (2.1)	0.011657 (2.4)
OWRK_8MI	5.722E-06 (2.7)	4.933E-06 (2.4)	5.729E-06 (2.8)	7.209E-06 (3.5)	7.811E-06 (3.9)	7.6E-06 (4.2)	6.2E-06 (3.0)	5.2E-06 (2.7)	4.63E-06 (2.7)

As the table above shows, the dollar effect of each variable as a component of the rent is roughly constant through time. That locational value is stable through time suggests that one need not be concerned about which point in a business cycle the evaluation occurs. This result is particularly intriguing since the period of analysis covered a substantial business cycle. The time period of discussion [1988-1996] covered a cycle that began at the tail end of a building boom, which quickly turned into a recession, and was then followed by a modest recovery. Because of the long-term nature of real estate investment, this is an important finding. It implies that the locational component of real estate decision making need not be affected by market timing.

The following charts track through time the overall change in rent, along with the change in the effect of its components. The lines of the graph below consist of the regression coefficients as applied to a prototypical office building. In each year, the equation was estimated separately, generating coefficients for each year as a separate model. Then, the resulting model for each year was used to predict the dollar contribution in rent of each variable. For instance, in the case of an inner region building, the ¼ mile neighborhood income contributes a fairly steady $3 to its overall rent, and is slightly different in each year. The vertical summation of those lines yields one of the overall rents shown above them, according to an inner or suburban location.

Analysis of Rent Through Time



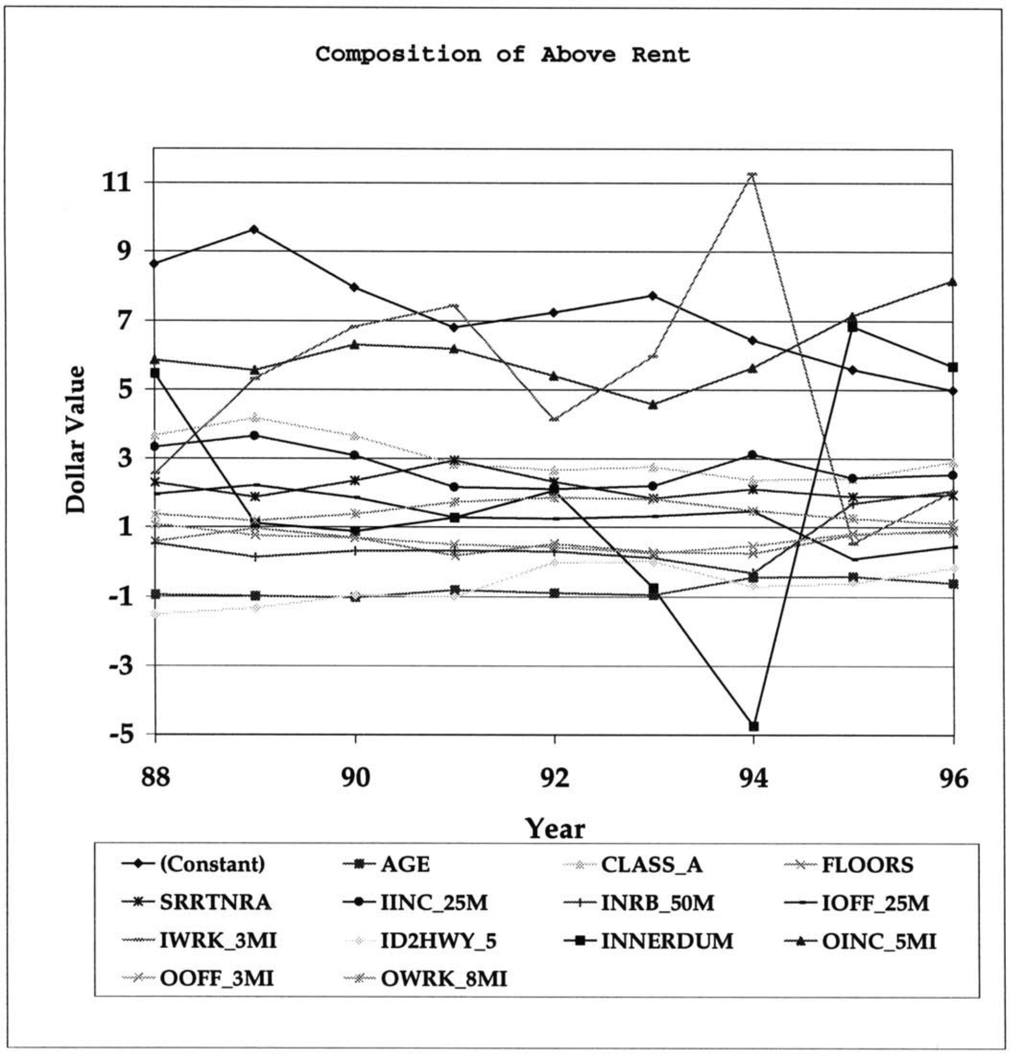


Figure 12

The interpretation of this phenomenon may be premature without studying the behavior of spatial variables through multiple economic cycles, but this single case implies that location is a fairly consistent contributor to the rent of office space. As data of this type and quality continue to accrue in commercial databases it will become possible to re-examine these questions concerning the relationships between space, time and the business cycle.

Model Prediction of Rent Potential

Using data from the metropolitan Boston area together with the spatial model, the GIS allowed the rapid calculation of this map below, showing a potential rent for *all* locations across the study space. Holding the building qualities as constant the map reveals the variation in rent across space. This can alternatively be called the locational rent or the spatial demand. This is a dramatic addition to standard regression analysis, with which it is possible to use the variable coefficients from the fitted model to predict the dependant variable. For example, using the base model of building characteristics only, one could predict that a two-year-old Class A building of 10 floors and 100,000 square feet would command $31.51 in rent. Using the spatial variables, the prediction is more realistically shown a function of its location.

Surface of Potential Rent

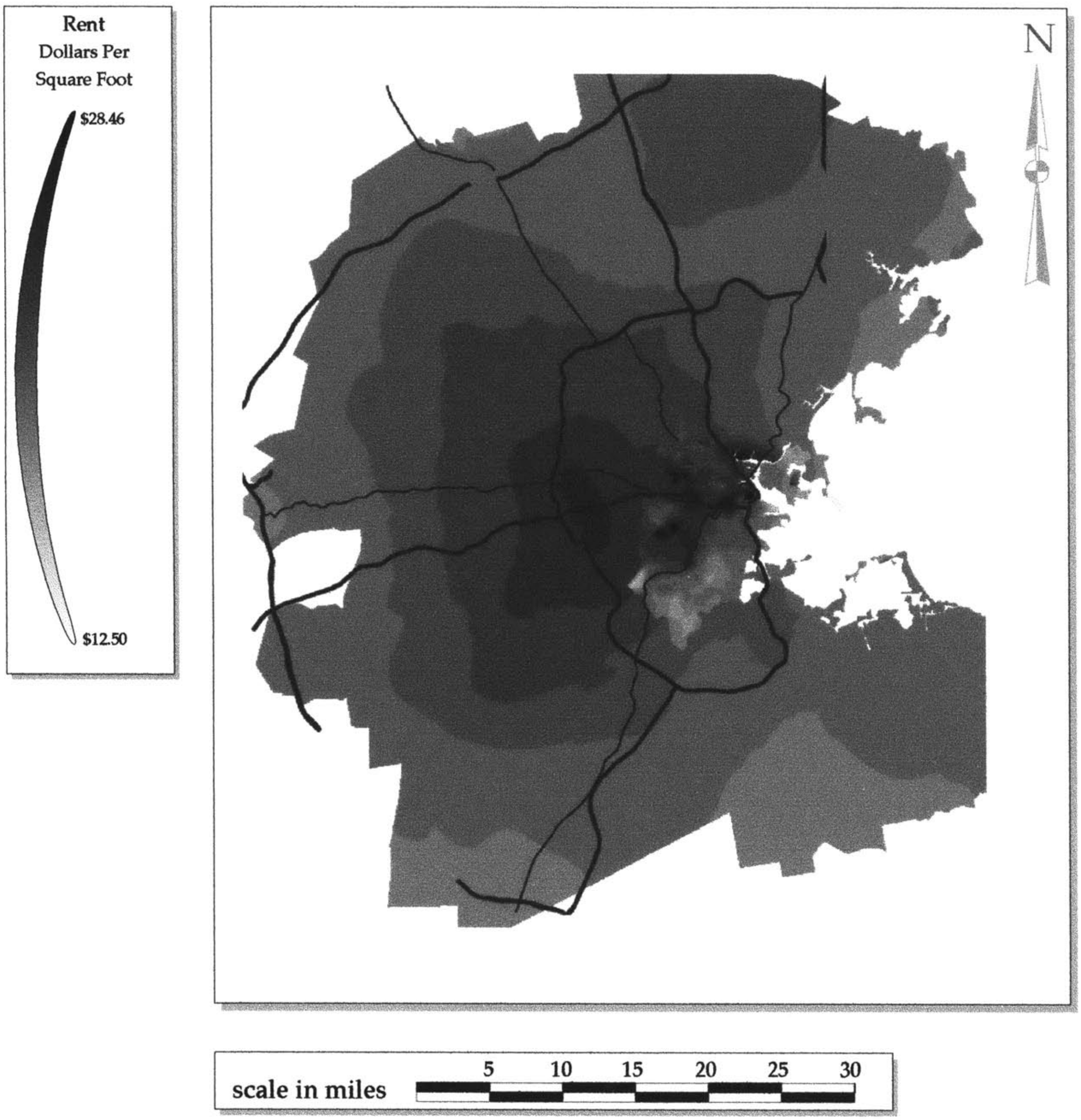


Figure 13.

A close look at this map of potential rent reveals several important things. First, the model does an excellent job of highlighting known hot spots of locational rent such as downtown Boston, Harvard Square, and Back Bay. The differentiation within the city of Boston is striking, yet not unlikely, as Boston has a very heterogeneous composition with respect to income levels, shopping centers, and office concentrations. Other areas of particular interest are the hot spots in Brookline. They are driven by the income of the specific neighborhoods in which they lie. Both rent peaks spots in Brookline are near the borders of both Boston (Brighton) and Newton. A developer in both of these bordering towns could take advantage of the value of the location played off against three different jurisdictions. The income levels of Boston are much lower, but the other locational factors are still strong, indicating that places exist which are proximate enough to derive some additional locational rent, yet exhibit lower costs. A model of potential rent for office space aids the planners and developers / investors of the region make more informed land use decisions.

Another engaging aspect of the map of spatial demand is the large island of high rents just east of Route 128 near the intersections with Route 9 and the Massachusetts Turnpike. It is worth adding here that this result was produced in the absence of any variable related to proximity to the highways themselves. This area just happens to have excellent access to a large number of workers where there is already some significant commercial development, and relatively high income. The towns covered by the island are Newton and Waltham, with Wellesley and Weston close-by. Furthermore, this island of higher rent is not an isolated phenomenon; it is surrounded by roughly concentric bands of lower rent except for the peaks within the inner region. This is consistent with a polycentric theory of urban growth and decentralizations [11]. Whereas downtown Boston was the economic center of the port economy of a century ago, the map of the model suggests that the intersection of the Massachusetts Turnpike and Route 128 is the economic center of the office market in an increasingly suburban, information-based economy.

Review of Residuals

Another feature of using a GIS is that, because any spatially referenced value can be displayed cartographically, it is possible to view the residuals as well as the prediction residuals. The following histogram of the residuals shows them to be normally distributed without any discernable trend. Upon visual inspection of the spatial distribution of the residuals, it is again difficult to detect any obvious pattern or bias. If there were a trend in either case, it would suggest a mis-specification error.

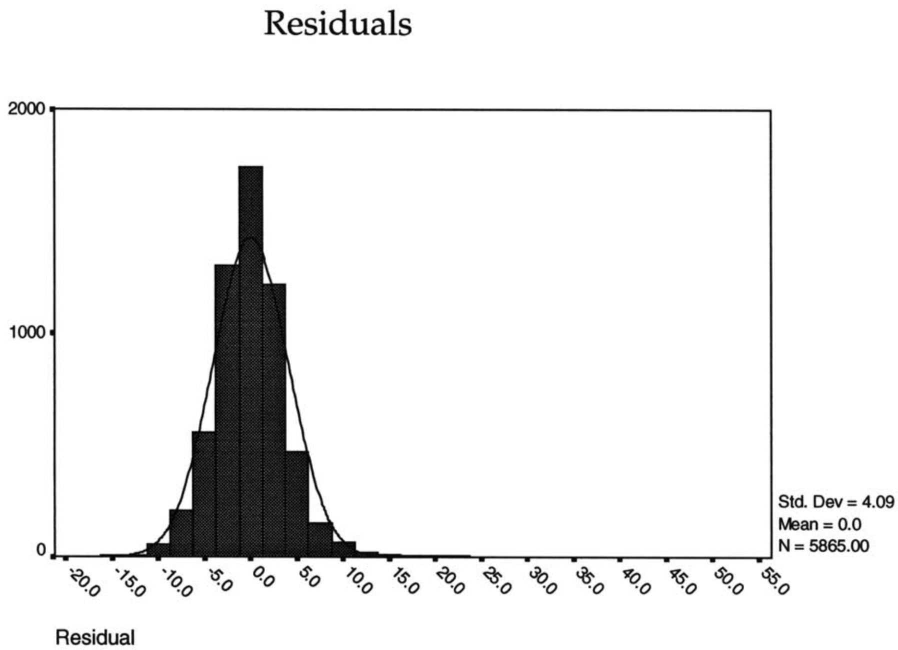


Figure 14.

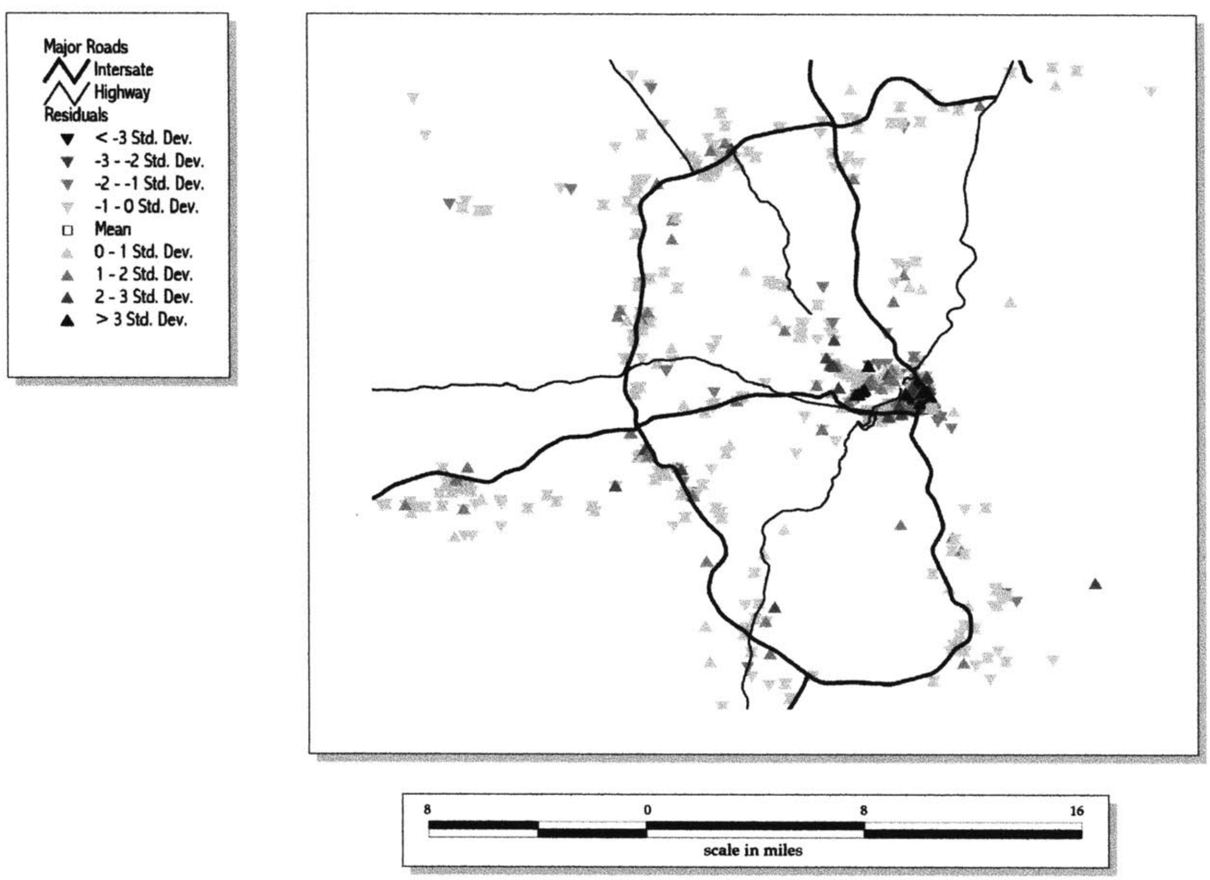


Figure 15.

CHAPTER FIVE
SUMMARY AND CONCLUSIONS

Chapter Introduction

The research conducted in this thesis served a three-fold purpose: it created a new model, predicted a surface of potential rent, and provided insight into urban economic theory. It created a spatially cognizant and, thereby more accurate, hedonic price model of the rent for office space, and in doing so calibrated a surface of locational rent for the entire study area. When combined with traditional non-spatial factors, the product is a surface of rent for an identical building as a function of its location. These outcomes create not only a locational price index for metropolitan Boston, but a methodology which can be applied to other markets, both substantive and geographical (such as housing markets in California). Before such additional research is conducted, some issues deserve additional attention. They include separating the dataset into two, the modifiable areal unit problem, and other naivetes of spatial modeling, the issue of stability over time, and finally, the probability of missing variables.

The Two Markets

A significant outcome of the process of modeling locational rent was the discovery that different spatial resolutions yielded different results based on location. The variation in rent was better captured with smaller radii in the inner region and larger radii in suburbs. The criteria for such a separation were briefly described in Chapter 3. What is important here is to acknowledge the model's overall sensitivity to this construction.

The data was split into two subsets because there are at least two drastically different styles of commercial development present in Boston. There is no question that the offices downtown are very different from their suburban counterparts such as those in Framingham. However, there is a considerable range of locational character between those two extremes. Creating a separate subset for each town would be an extreme, anathema to the assumptions of the overall thesis that the richness of location can be quantified and not just treated as residual value. To more appropriately divide the data into several subsets, it would be necessary to employ sophisticated spatial analysis techniques, which were not within the scope of this thesis. However, the main difference is captured when small radii are used for the inner region and larger radii for the suburbs. When examining the inner region, one should consider walking distance as the main spatially restrictive criterion. When evaluating property in the suburbs, it is driving times and distances which are relevant.

Spatial Analysis

A different, yet related problem to the definition of subsets is what is referred to in the geography literature as the "modifiable areal unit problem" [4]. The areal unit problem is the dilemma caused by the fact that, in this type of analysis, the results are sensitive to the variation in the spatial construction of the variable. In other words, the results of this model are dependant on both cell size and radii or resolution of the spatial factors. It was theoretically but not practically possible in this research to discover the optimal size, resolutions, or functional forms.

While this analysis is a unique application of GIS technology to econometric research using private, disaggregate data in a public setting, it is still spatially naïve. The explicit quantification of location unmistakably added explanatory power to the pricing model (roughly 10%), and is an improvement over previous such studies which either neglect space altogether or treat it very cursorily. However, the remaining naïveté is that it ignores underlying spatial structure present in data. The result of which is that the error may (and probably does) suffer from auto-correlation. There are techniques in spatial statistics that can explicitly test for and model this behavior, which were not applied here. The result of correcting for the autocorrelation of the error terms would be mixed. The model would be improved, due to the enhanced efficiency of the estimators, but the overall fit may be diminished [4, 5].

The principle aim of the thesis was to quantify and price location, and this was accomplished to a certain degree with the metrics tested {distance based, agglomerative, amenity oriented}. However, there is still much richness of place that is unexplained. The challenge in better capturing and quantifying place is finding or collecting the appropriate data, and then discovering the scales at which the spatial process is revealed.

Missing Variables

With respect to specifically modeling the office market there are several variables, which were of great interest, but remain untested. One basic element of the character of location, which is important in terms of both amenity and agglomerative benefits, is the degree to which a given area is walkable. By separating the dataset into the inner cities, in which resolutions based on walking were used, and the suburbs, where they were based upon driving, I assume that there are not walkable or urban style suburban office areas. It would be better to have a metric in the model that could capture this effect. This analysis tried to capture this effect through road network density. It was a calculation of the number of road segments within a certain radius of a point. Although it implied a commercial urban density, none of the tested variations in size proved significant. Another locational variable that was not evaluated was transportation capacity. Proximity to transportation infrastructure proved an inadequate proxy, and thus more detailed data is needed, such as travel time on the street network from office buildings to highway access points. The retail amenity measure also suffered from a lack of detail. It only included shopping centers and did not take into account nearby restaurants, convenience stores or business services such as copy centers or dry cleaners. Small, independent retailers of the kind that are particularly common in Boston were not included yet are considered quite important assets. Other variables might include the view or view-shed of a particular building, which encompasses issues of view and civic architecture. The list of potential locational variables is long and could possibly be exhausted given more time and resources.

Concluding Remarks

This thesis explored one possible way of improving upon real estate pricing for the purpose of economic efficiency. The contributions of the thesis, other than the final model, may prove to be of longer lasting influence. They include the proposed methodology of recalculating spatial variables into repeatable units across markets, geographies, and times, enabling a more cumulative approach to the understanding of spatial phenomena in hedonic price modeling. Also of significance is the practice of developing spatially context sensitive explanatory variables, which in this case led to the dividing of the dataset into two categories. More research is needed in the systemization of breaking metropolitan areas into submarkets, be it for commercial office space or housing. Ultimately in this research, it meant the difference between markets in which the dominant mode of transportation was walking or driving, but other metropolitan markets may have other spatial differentiates. Lastly and perhaps most dramatic is the ability through the use of GIS as an analytic component to econometric study to predict a spatially continuous surface of potential office rent. The thesis makes considerable inroads in the development of spatially sensitive hedonic models, but more research is required in the development and use of such techniques to further resolve the key challenge of this thesis regarding the accurate and efficient pricing of location.

BIBLIOGRAPHY

1. Alonso, W., Location and Land Use, Harvard University Press., Cambridge, MA (1964).

2. Anselin, L., Florax, J.G.M., New Directions in Spatial Econometrics, Springer, (1995).

3. Bailey, M. J., Muth, R. F., Nourse, H. O., A Regression Method for Real Estate Price Index Construction, *Journal of the American Statistical Association*, **58** (1963).

4. Bailey, T. C., and Gatrell, A., Interactive Spatial Data Analysis, Longman Scientific and Technical, Essex, England (1995).

5. Burrough, P. A., Principles of Geographical Information Systems for Land Resources Assessment., Oxford University Press, Inc., New York, NY. (1986).

6. Brown, J., Rosen, H. S., On the Estimation of Structural Hedonic Price Models, *Econometrica*, **50** (3) (1982).

7. Can, A., The Measurement of Neighborhood Dynamics in Urban House Prices, *Economic Geography*, **66** (3) (1990).

8. Can, A., Megbolugbe, I., Spatial Dependance and House Price Index Construction, *Journal of*
Real Estate Finance and Economics, **14** (1/2) (1997).

9. Central Transportation Planning Staff, Ten Park Plaza, Suite 2150 Boston, Massachusetts 02116.

10. Clapp, J., Pollakowski, H., Lynford, L., Intrametropolitan Location and Office Market Dynamics, *Journal of the American Real Estate and Urban Economics Association*, **20** (1) (1992).

11. DiPasquale, D. and Wheaton, W.C., Urban Economics and Real Estate Markets, Prentice Hall,
Englewood Cliffs, NJ (1996).

12. Geoghega, J, Wainger, L, Bockstael, N., Spatial Landscape Indicies in a Hedonic Framework:
An Ecological Analysis Using GIS, *Ecological Economics*, **23** (3) (1997).

13. Goetzmann, W., Spiegel, M., A Spatial Model of Housing Returns and Neighborhood Substitutability, *Journal of Real Estate Finance and Economics*, **14** (1/2) (1997).

14. Helsly and Sullivan, Urban Subcenter Formation, Regional Science and Urban Economics, **21** (2) (1991).

15. Kennedy, P., A Guide to Econometrics, MIT Press, Cambridge, MA (1992).

16. McDonald, J.F., The Identification of Urban Employment Subcenters, *Journal of Urban Economics*, **38** (1) (1995).

17. Mills, E., The Value of Urban Land, The Quality of the Urban Environment, Resources for the
Future, Washington, DC., (1969).

18. Mills, E., Office Rent Determinants in the Chicago Area, *Journal of the American Real Estate and*
Urban Economics Association, **20** (1) (1992).

19. Mills, E., and Hamilton, Bruce, Urban Economics, Harper Collins, New York, NY (1994).

20. Moses, L. N., Toward a Theory of Intra-Urban Wage Differentials and Their Influence on Travel Patterns., Papers of the Regional Science Association, **9** (1962).

21. Mun, S., and Hutchinson, B., Empirical Analysis of Office Rent and Agglomeration Economies: A Case Study of Toronto, *Journal of Regional Science*, **35** (3) (1995).

22. Muth, R., Cities and Housing, University of Chicago Press, Chicago., (1969).

23. NRB Dataset, National Research Bureau, 45 Danbury Road, Wilton, CT 06897.

23. Pindyck, R.S., and Rubenfeld, D.L., Econometric Models & Economic Forcasts, McGraw Hill, NewYork, NY (1991).

25. Ricardo, D., Principles of Political Economy and Taxation, 3rd ed. John Murray, London., (1886).

26. Rosen, S., Hedonic Price and Implicit Markets: Productive Differentiation in Pure Capitalism,
Journal of Political Economy, **82** (1974).

27. Samuelson, P., Economics, McGraw-Hill, New York, NY., (1980).

28. Sivitanidou, R., Do Office-Commercial Firms Value Access to Service Employment Centers:
A Hedonic Value Analysis within Polycentric Los Angeles, *Journal of Urban Economics*, **40** (2) (1996).

29. Sivitanidou, R., Urban Spatial Variation in Office-Commercial Rents: The Role of Spatial Amenities and Commercial Zoning, *Journal of Urban Economics*, **38**(1) (1995).

30. Shukla, and Waddel, Firm Location and Land Use in Discrete Urban Space, *Regional Science and Urban Economics*, **21** (2) (1991).

31. Steele, M., and Goy, R., Short Holds, the Distributions of First and Second Sales, and Bias in the Repeat-Sales Price Index, Journal of Real Estate Finance and Economics, 14: 133-154 (1997).

32. Timothy, D., Intraurban Wage Differentials and Commuting Time. Ph. D. Dissertation. (1994).

33. TWR, Torto Wheaton Research, CB Commerical, Boston, MA.

34. U.S. Bureau of the Census, Washington, DC 20233.

35. Waddell, P., Berry, B., Hoch, I., Residential Property Values in a Multinodal Urban Area: New Evidence on the Implicit Price of Location, *Journal of Real Estate Finanace and Economics*, **7** (2) (1993).

36. Wheaton, W. C., and Torto, R. G., Office Rent Indicies and Their Behavior over Time, *Journal of Urban Economics*, **35** (2) (1994).

37. Von Thunen, Isolated State, Pergamon Press, New York, N.Y. (1966).

APPENDIX A

DATA DICTIONARY

VARIABLES IN EQUATION

Variable	Description
AGE	Lessor of the Number of Years since a building was built or renovated
CLASS_A	A Categorical Variable for Class A Space [1/0]
FLOORS	Number of floors
SRRTNRA	Square Root of the Net Rentable Area
RENT	Average asking rent of a building
ID2HWY_5	Square Root of the Distance from to Nearest Highway
IINC_25M	Average Median Block Group Income within ¼ Mile
INRB_50M	Number of Shopping Centers within ½ Mile
IOFF_25M	Number of Office Buildings within ¼ Mile
IWRK_3MI	Number of Office Workers within 3 Miles
INNERDUM	A Categorical Variable for Inner Region Buildings
OINC_5MI	Average Median Block Group Income within 5 Miles
OOFF_3MI	Number of Office Buildings within 3 Miles
OWRK_8MI	Number of Workers within 8 Mikes

WORKERS

STF-3A Census Variable	Description
P0780001	Executive, Administrative, Managerial
P0780002	Professional Specialty
P0770010	Finance, Insurance, Real Estate
P0770016	Other Professional and Related Services
P0770017	Public Administration

APPENDIX B

DESCRIPTIVE STATISTICS OF DATASETS

TABLE 1: 1990 CENSUS DEMOGRAPHICS

Variable	Mean	Std Dev	Range	Sum	Valid N
POP100__	1100.25	645.59	5604	3509803	3190
MEDHHINC	41697.56	19045.06	150001	1.33E+08	3190
ADMIN_SU	103.19	73.69	804	329187	3190
EX_ADMIN	93.74	82.07	766	299046	3190
FIRE	52.08	43.66	414	166145	3190
O_PROFESS	55.49	48.55	456	177026	3190
PROF_SPEC	108.34	94.93	775	345592	3190
PUB_ADMIN	25.5	22.13	244	81332	3190
WORKERS	438.35	306.43	2541	1398328	3190

TABLE 2: TWR OFFICE MARKET DATA

BUILDINGS WITHING STUDY AREA WITH
AT LEAST ONE VALID RENT OBSERVATION

Variable	Mean	Std Dev	Range	Minimum	Maximum	Sum	Valid N
BUILT	1950	106.48	1992	1	1993	1928552	989
CLASS_A	0.17	0.38	1	0	1	171	989
FLOORS	5.96	6.03	59	1	60	5899	989
NRA	101550.8	155743.8	1593533	4000	1597533	1E+08	989
RENOV	630.13	924.36	1996	0	1996	623201	989

APPENDIX C

DESCRIPTIVE STATISTICS OF VARIABLES IN EQUATIONS

TABLE 1: ALL CASES

Variable	Mean	Variance	Range	Minimum	Maximum	Sum	Valid N
AGE	18.74	295.48	142	2	144	109898	5865
CLASS_A	0.2	0.16	1	0	1	1200	5865
FLOORS	6.58	44.61	59	1	60	38599	5865
SRRTNRA	290.88	27618.65	1200.69	63.24555	1263.9355	1706036.52	5865
RENT	19.36	34.48	72	3	75	113560.64	5865
ID2HWY_5	9.2	141.55	50.8	0	50.80057	53964.11	5865
IINC_25M	16840.23	462488371	103352.77	0	103352.77	98767971.08	5865
INRB_50M	1.23	2.84	6	0	6	7208	5865
IOFF_25M	32.2	2235.11	164	0	164	188877	5865
IWRK_3MI	77350.18	6665561633	224098.19	0	224098.19	453658824.5	5865
INNERDUM	0.49	0.25	1	0	1	2856	5865
OINC_5MI	30122.12	923175031	80007.32	0	80007.32	176666229.3	5865
OOFF_3MI	38.48	2591.7	675	0	675	225675	5865
OWRK_8MI	123070.6	2.32E+10	586232.69	0	586232.69	721809040.9	5865

TABLE 2: INNER REGION ONLY

Variable	Mean	Variance	Range	Minimum	Maximum	Sum	Valid N
AGE	21.3	454.01	141	3	144	60837	2856
CLASS_A	0.23	0.18	1	0	1	668	2856
FLOORS	9.72	70.25	59	1	60	27754	2856
SRRTNRA	324.76	43963.71	1174.49	89.44272	1263.9355	927523.97	2856
RENT	21.34	47.36	72	3	75	60953.08	2856
ID2HWY_5	18.89	107.5	50.8	0	50.80057	53964.11	2856
IINC_25M	34582.62	336130308	100904.24	2448.5293	103352.77	98767971.08	2856
INRB_50M	2.52	2.57	6	0	6	7208	2856
IOFF_25M	66.13	2346.14	163	1	164	188877	2856
IWRK_3MI	158844.13	741298431	177173.56	46924.625	224098.19	453658824.5	2856
INNERDUM	1	0	0	1	1	2856	2856
OINC_5MI	0	0	0	0	0	0	2856
OOFF_3MI	0	0	0	0	0	0	2856
OWRK_8MI	0	0	0	0	0	0	2856

TABLE 3: SUBURBAN REGION ONLY

Variable	Mean	Variance	Range	Minimum	Maximum	Sum	Valid N
AGE	16.3	132.95	97	2	99	49061	3009
CLASS_A	0.18	0.15	1	0	1	532	3009
FLOORS	3.6	2.09	13	1	14	10845	3009
SRRTNRA	258.73	9989.99	664.77	63.24555	728.01099	778512.55	3009
RENT	17.48	15.01	23.75	5	28.75	52607.56	3009
ID2HWY_5	0	0	0	0	0	0	3009
IINC_25M	0	0	0	0	0	0	3009
INRB_50M	0	0	0	0	0	0	3009
IOFF_25M	0	0	0	0	0	0	3009
IWRK_3MI	0	0	0	0	0	0	3009
INNERDUM	0	0	0	0	0	0	3009
OINC_5MI	58712.61	120520285	48181.16	31826.16	80007.32	176666229.3	3009
OOFF_3MI	75	2312.39	674	1	675	225675	3009
OWRK_8MI	239883.36	1.72E+10	575873.1	10359.589	586232.69	721809040.9	3009